A Fixed Rate Speech Coder Based on the Filter Bank Method with Non-Uniform Bandwidths

Byeong-Gwan Iem

Department of Electronic Engineering, Gangneung-Wonju National University, Gangneung, Korea

Ijfis

Abstract

A fixed rate speech coder exploiting the filter bank with non-uniform bandwidth is proposed. The speech coder is based on the inflection point (IP) detection. The speech signal is filtered by a group of bandpass filters with non-uniform bandwidths. Each bandpass filtered signal is processed by the IP detector, and the obtained IP pattern is compared with the entries of the IP pattern database. To facilitate the database search, the IP pattern entries are grouped into two kinds: IP patterns for voiced and the unvoiced speech. Before the search, the obtained IP patterns are checked if they belong to the voiced or the unvoiced. For that purpose, the short-time energy of IP pattern from the first band is calculated and compared to the predetermined threshold value. In addition to the unvoiced/voiced decision, the non-uniform bandwidth bandpass filters are used for the filter bank to speed up the database search. By using the non-uniformly spaced bandpass filters, the number of filters in the filter bank becomes smaller. Transmitted information for each subband includes the IP pattern address of the database and amplitude of each IP pattern. In the receiver, the decoder recovers the subband signals using the received addresses and the energy information, and reconstructs the speech via the filter bank summation. Through computer simulation, the usefulness of the proposed technique is confirmed. The SNR performance of the proposed method is approximately 9 dB with relatively low bitrate of 12.8 kbps.

Keywords: Non-uniformly spaced filter bank, Non-uniform sampling, Inflection point detection

Received: Dec. 4, 2018
 Revised : Jun. 7, 2019
 Accepted: Jun. 24, 2019

Correspondence to: Byeong-Gwan Iem
 (ibg@gwnu.ac.kr)
 ©The Korean Institute of Intelligent Systems

©This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

1. Introduction

Non-uniform sampling based speech coding has been studied last 20 years [1–9]. Since it is non-uniform, the coding result shows variable data rate [1–7]. As a remedy for this problem, the filter bank method had been proposed [8, 9]. Speech signal is preprocessed with a filter bank, and the output of each bandpass filter is put into the inflection point (IP) detector. The obtained IP patterns are compared with the entries of IP database. Thus, the information to be transmitted or stored is the address of the database, and the data rate becomes constant rather than variable [8, 9]. In the decoder, the speech is reconstructed through interpolation using the received address and the IP database.

However, in the coding process, enormous computation may be required due to the database search. Thus in this paper, a new approach is proposed to reduce the amount of computation.

Figure 1. Enlarged plot of a speech signal with various inflection points [8].

That is, the non-uniformly spaced bandpass filters are selected for the filter bank. Contrary to the uniformly spaced filter bank method, the non-uniformly spaced filter bank can have the smaller number of bandpass filters. As results, the amount of computation can be reduced. In this paper, the bandwidths of the neighboring bandpass filters becomes twice as the center frequency goes high. In addition, to facilitate the search speed of the IP database, the voiced/unvoiced detection is performed using the short-time statistics as in [9]. The difference is to use only the first bandpass filter output rather than the whole output of the filter bank to save the computation. Since the first band of speech includes important information such as the fundamental frequency of the speech, the IP pattern from the first bandpass filter output is sufficient to get a statistics for the unvoiced/voiced decision. The paper is written as follows. In the second section, the definition of the IP and its detection algorithm are reviewed briefly. In Section 3, the structure of the proposed speech coder is explained in detail. In the next section, the short-time statistics for the unvoiced/voiced decision is considered. In Section 5, simulation results are provided to show the usefulness of the proposed method. In this section, the frequency response of non-uniformly spaced filter bank is also given.

2. Definition of Inflection Points and Their Detection

The inflection points include sample points where the slope changes happen as in Figure 1. They are points with simple slope changes (point c), and points at local maxima (point b) and local minima (point a). The inflection points are detected using following algorithm [7–9].

For consecutive samples x_1 , x_2 , and x_3 , a local maxima and minima point x_2 is obtained if the product of the consecutive

differences is less than 0, i.e.,

$$d_{21} \cdot d_{32} < 0, \quad (1)$$

where the differences of consecutive samples are defined as

$$d_{21} = x_2 - x_1, \quad (2)$$

$$d_{32} = x_3 - x_2. \quad (3)$$

The sample point with simple slope change is checked by following measure [8]:

$$\text{identifier (ID)} = \frac{|d_{21} - d_{32}|}{|d_{21}| + |d_{32}|}. \quad (4)$$

The range of identifier value is $0 < ID \leq 1$. If the ID value is above a predetermined threshold value, the sample point is decided to be an IP. The IP detection process is depicted in Figure 2.

3. The Structure of the Speech Coder

The structure of the speech coder is shown in Figure 3. A block of speech is bandpass filtered first using a filter bank of M bandpass filters. The bandwidths of the bandpass filters are non-uniformly spaced to reduce the number of filters. M

Figure 2. Inflection point detection algorithm [8].

Figure 3. Structure of the encoder/decoder.

filter outputs are then placed into the IP detector, and M IP patterns are obtained. The IP patterns are normalized, and compared with the entries of the IP pattern database to get the addresses of closest IP entries. In this process, the short-time energy of the IP pattern of the first band is used to decide if the speech block is unvoiced or voiced. The decision can facilitate the database search procedure. The information transmitted is the addresses of the closest IP entries and amplitudes of the obtained IP patterns.

4. The Short-Time Statistics of the Inflection Points for the Voiced/Unvoiced Decision

In the previous work [9], the IP rate and the short-time energies of IP patterns from all the subbands had been used for the voiced/unvoiced decision. In this paper, the short-time energy of the IP pattern from the first subband is used for the following reason. This reduces the amount of computation by processing the IP pattern from only one band. Furthermore, since the first bandpass filter acts as a low pass filter removing high frequency components, the output may include the significant fundamental frequency if the speech is voiced. As results, the IP patterns showing high short-time energy can be decided as the voiced part. Figure 4 shows simulation results of the short-time energy of uniformly sampled speech and that of the IP pattern from the first bandpass filter. The short time energy of the IP can be used to decide if a speech block is voiced.

Figure 4. Comparison of the short-time energy: (a) original speech signal, (b) short-time energy of the speech, and (c) the short-time energy of the IP of the first subband of the filter bank output.

5. Simulation Results

The computer simulation shows the usefulness of the proposed speech coding technique. The sampling frequency of a speech is 10 kHz, and the speech is segmented into 10 ms blocks with 50% overlapping. The number of bandpass filters in the non-uniformly spaced filter bank is 4. The center frequencies and bandwidths of the bandpass filters are (0, 660, 1650, 3630 Hz) and (330, 660, 1320, 2640 Hz), respectively. The frequency

response of the bandpass filters is shown in Figure 5(b). In Figure 5(a), the frequency response of the uniformly spaced bandpass filters is also given for comparison. The first band in Figure 5(b) includes the frequency band where the fundamental frequency of voiced speech resides [10–12]. Thus, the IP pattern from this band can be used for the voiced/unvoiced decision as pointed out in the previous section. The IP pattern database has 250 entries for each band, so the number of bits for the address of each IP is $\lceil \log_2 250 \rceil = 8$, where $\lceil x \rceil$ is the nearest integer greater than x . And the number of bits for the energy of an IP pattern is 8 bits. For each bandpass filter output, 16 bits are required for the address and amplitude. And there are 200 10 ms speech blocks in a second. Since 4 non-uniformly spaced bandpass filters are used for the filter bank, the data rate is $(16 * 200 * 4)$ bits/second = 12.8 kbps. Since the proposed method uses non-uniformly spaced bandpass filters, it needs a smaller number of filters compared to the previous method which used the uniformly spaced bandpass filters [8, 9]. In this simulation test, the proposed method has 4 bandpass filters as shown in Figure 5(b) whereas the existing one exploits 10 uniformly spaced filters as in Figure 5(a). Thus, it is expected that the proposed method takes about 40% of the processing time of the previous method. When the processing time was measured using ‘tic’ and ‘toc’ command in MATLAB simulation, the proposed method showed 133 seconds while the previous one took 348 seconds. Figure 6 shows the processed signal using the proposed method. Figure 6(a) is the original signal, and Figure 6(b) is the reconstructed signal at the receiver. From Figure 6(b), the usefulness of the proposed method can be shown. The signal-to-noise (SNR) of the proposed speech coder is calculated as follows:

$$\text{SNR} = 10 \log_{10} \left[\frac{\text{signal power}}{\text{noise power}} \right], \quad (5)$$

where the noise is the difference between the original and the reconstructed signal. The SNR value obtained is 9.03 dB.

6. Conclusion

A new speech coding technique using the non-uniformly spaced filter bank method has been proposed. The coding technique is a non-uniform sampling method based on the IP detection. Unlike existing IP detection based coding methods with the filter bank, the proposed coder exploits the filter bank of which bandpass filters show the non-uniform and increasing bandwidths along with non-uniformly spaced center frequencies. Resulting effects

Figure 5. Frequency response of the bandpass filters of the filter bank: (a) filter bank with uniformly spaced filters and (b) filter bank with non-uniformly spaced filters.

Figure 6. Processing results of the IPD based coding: (a) original speech and (b) reconstructed speech.

are the smaller number of bandpass filters in the filter bank and the reduction in computation. The simulation result shows that the proposed technique produces relatively low bitrate in encoding and the reconstructed signal closely resembles to the original signal with SNR of 9.03 dB.

Conflict of Interest

No potential conflict of interest relevant to this article was reported.

Acknowledgments

This study was supported by Gangneung-Wonju National University.

References

- [1] M. Bae, W. Lee, and D. Kim, "On a new vocoder technique by the nonuniform sampling," in *Proceedings of IEEE Military Communications Conference*, McLean, VA, 1996, pp. 649-652. <http://doi.org/10.1109/milcom.1996.569428>
- [2] M. Budaes and L. Goras, "On speech signals reconstruction from local extreme values," in *Proceedings of International Symposium on Signals, Circuits and Systems*, Iasi, Romania, 2005, pp. 315-318. <http://doi.org/10.1109/ISSCS.2005.1509917>
- [3] L. Davisson, "Data compression using straight line interpolation," *IEEE Transactions on Information Theory*, vol. 14, no. 3, pp. 390-394, 1968. <http://doi.org/10.1109/TIT.1968.1054160>
- [4] J. Mark and T. Todd, "A nonuniform sampling approach to data compression," *IEEE Transactions on Communications*, vol. 29, no. 1, pp. 24-32, 1981. <http://doi.org/10.1109/TCOM.1981.1094872>
- [5] B. G. Iem, "A nonuniform sampling technique based on inflection point detection and its application to speech coding," *The Journal of the Acoustical Society of America*, vol. 136, no. 2, pp. 903-909, 2014. <https://doi.org/10.1121/1.4884882>
- [6] B. G. Iem, "A nonuniform sampling technique and its application to speech coding," *Journal of Korean Institute of Intelligent Systems*, vol. 24, no. 1, pp. 28-32, 2014. <https://doi.org/10.5391/JKIIS.2014.24.1.028>
- [7] B. G. Iem, "A low bit rate speech coder based on the inflection point detection," *International Journal of Fuzzy Logic and Intelligent Systems*, vol. 15, no. 4, pp. 300-304, 2015. <https://doi.org/10.5391/IJFIS.2015.15.4.300>
- [8] B. G. Iem, "A fixed rate speech coder based on the filter bank method and the inflection point detection," *International Journal of Fuzzy Logic and Intelligent Systems*, vol. 16, no. 4, pp. 276-280, 2016. <https://doi.org/10.5391/IJFIS.2016.16.4.276>
- [9] B. G. Iem, "A fixed rate speech coder with improved database search," *International Journal of Fuzzy Logic and Intelligent Systems*, vol. 17, no. 4, pp. 289-293, 2017. <https://doi.org/10.5391/ijfis.2017.17.4.289>
- [10] L. Rabiner and R. Schafer, *Digital Processing of Speech Signals*. Englewood Cliffs, NJ: Prentice-Hall, 1978.
- [11] T. F. Quatieri, *Discrete-Time Speech Signal Processing*. Upper Saddle River, NJ: Prentice-Hall, 2002.
- [12] G. Lee and W. G. Kim, "Emotion recognition using pitch parameters of speech," *Journal of Korean Institute of Intelligent Systems*, vol. 25, no. 3, pp. 272-278, 2015. <https://doi.org/10.5391/jkiis.2015.25.3.272>

Byeong-Gwan Iem received his B.S. and M.S. from Yonsei University, Seoul, Korea, in 1988 and 1990, respectively. He received his Ph.D. from the University of Rhode Island, RI, USA in 1998. He is a professor

at Gangneung-Wonju National University, Gangneung, Korea. His areas of study interests are DSP and its applications.

E-mail: ibg@gwnu.ac.kr